Examen (durée : 1h30)

Documents autorisés : feuille A4 recto verso manuscrite.

1 Exercice I [Classification]

Soit les données d'apprentissage suivantes :

$\{(x_i, y_i)\}_{i=1}^7 = \{(-3, 1), (-2, 1), (-1, -1), (0, -1), (1, -1), (2, 1), (3, 1)\}$, avec $(x_i, y_i) \in \mathbb{R}\times\{-1, 1\}$.

1. Représentez les données d'entrée sur un graphique en utilisant le symbole $\times$ pour les données de classe 1 et le symbole $\circ$ pour les données de classe -1. Est-ce que ces données sont linéairement séparables ?
2. Soit la fonction $g : \mathbb{R} \to \mathbb{R}^2$ telle que $g(x) = (x, x^2)$ permettant de représenter les données x_i dans $\mathbb{R}$ par leurs projections z_i dans $\mathbb{R}^2$. Dessinez sur un graphique les données $\{z_i = g(x_i)\}_{i=1}^n$ en utilisant les mêmes symboles décrits dans 1). Est-ce que les représentations z_i sont linéairement séparables ?
3. Proposez un algorithme d'apprentissage permettant de séparer linéairement les représentations z_i. Quelle est la complexité en temps de cet algorithme ?
4. Donner un noyau reproduisant k associé au mapping g et déduire un algorithme d'apprentissage adapté pour classer les données x_i sans utiliser les représentations z_i. Quelle est la complexité en temps de cet algorithme ?
5. Entre les deux algorithmes, lequel recommanderiez-vous ? Pourquoi ?

Exercice II [Noyaux]

1. On dispose d'un jeu de données $S = \{x_1, \dots, x_n\}$ et d'un noyau $k(\cdot, \cdot)$. Soit K la matrice de Gram du noyau k, i.e., $K_{ij} = k(x_i, x_j)$, $i, j = 1, \dots, n$. Montrez que s'il existe un mapping $\Phi(\cdot)$ tel que $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j)\rangle$ alors la matrice K est semi-définie positive (sdp).
2. Montrez l'inverse de 1., c'est à dire donnez une fonction $\Phi(\cdot)$ telle que si K est sdp alors $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j)\rangle$.
3. On considère le noyau *différence* défini par $k(x, x') = \|x - x'\|$. Montrer que ce noyau n'est pas sdp.
4. Donnez deux inconvénients des méthodes à noyaux. Comment y remédier ?

Exercice III [Complexité de Rademacher]

La complexité empirique de Rademacher d'une classe de fonctions $\mathcal{H}$ est définie comme :

$$\hat{R}_S(\mathcal{H}) = \frac{1}{n} E_\sigma \left[\sup_{h \in \mathcal{H}} \sum_{i=1}^{n} \sigma_i h(x_i) \right],$$

avec $x_i \in \mathcal{X}$, $i = 1, \ldots, n$, des observations et σ_i, $i = 1, \ldots, n$, des variables aléatoires indépendantes, tirées selon la loi de Rademacher (i.e., $\mathbb{P}(\sigma_i = +1) = \mathbb{P}(\sigma_i = -1) = 1/2$).

1. Quelle est la complexité de Rademacher d'un espace d'hypothèse réduit à une seule hypothèse ?
2. Soit $\mathcal{H}$ la famille de fonctions définie sur $\mathcal{X}$ comme suit : $\mathcal{H} = \{x \mapsto 1_{x=z} : z \in \mathcal{X}\}$, avec $1_{x=z} = 1$ si $x = z$ et 0 sinon.
   (a) Supposons que l'ensemble $S = \{x_i\}_{i=1}^n$ contient que des observations distinctes. Montrez que $\hat{R}_S(\mathcal{H}) \leq \frac{1}{n}$.
   (b) Supposons que l'ensemble $S = \{x_i\}_{i=1}^n$ contient une seule observation répétée n fois. Montrez que $\hat{R}_S(\mathcal{H}) \leq \frac{1}{n} E_\sigma \left[\left| \sum_{i=1}^{n} \sigma_i \right| \right]$.
   (c) En déduire, utilisant l'inégalité de Jensen, que $\hat{R}_S(\mathcal{H}) \leq \frac{1}{\sqrt{n}}$.